Lessons Learned Bringing Up Early Adopter PCIe® 4.0 Links

Richard Solomon

Synopsys

Disclaimer

<u>Presentation Disclaimer</u>: All opinions, judgments, recommendations, etc. that are presented herein are the opinions of the presenter of the material and do not necessarily reflect the opinions of the PCI-SIG®.

Acknowledgements

- **Many thanks to Torrey Lewis for actually DOING the integration testing and providing screen shots and other data for the presentation!**

Agenda

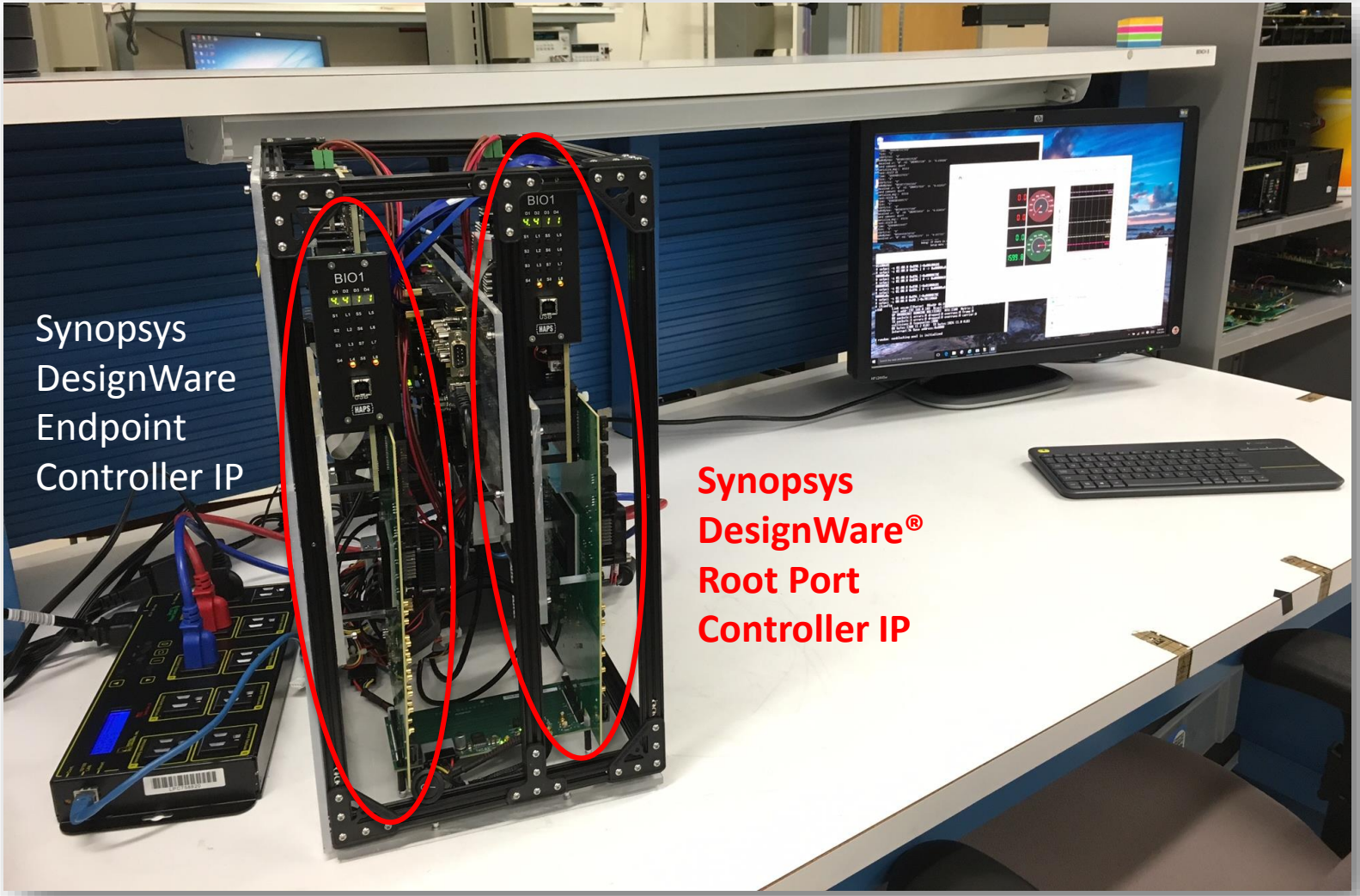
- **Overview**
- **Testing Setup (Hardware/Software)**
- **Interesting PCIe 4.0 Features**
- **Failure Manifestations**
- **Resolution**

Testing Setup

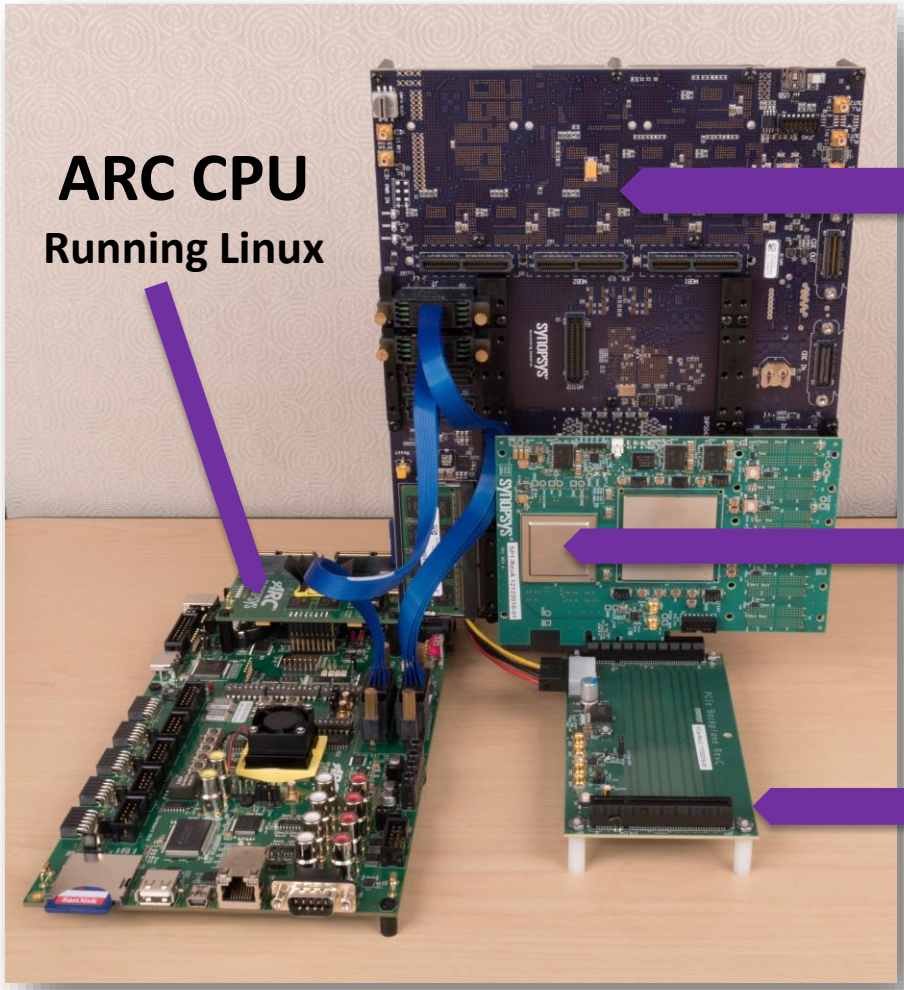
Testing Setup (Hardware/Software)

- **Synopsys PCIe 4.0 Root Complex**
  - Linux Running on ARC CPU single-board computer
  - Includes “RAS-DES” features
- **Synopsys PCIe 4.0 Endpoint**
  - Dedicated demo design with DMA engines, data movers, comparators, etc. to stimulate and respond to PCIe traffic
  - Includes “RAS-DES” features
- **Both on Synopsys HAPS Prototyping Platform**
  - FPGA-based
  - Includes Generic Embedded Logic Analyzer

Test Setup – Overview

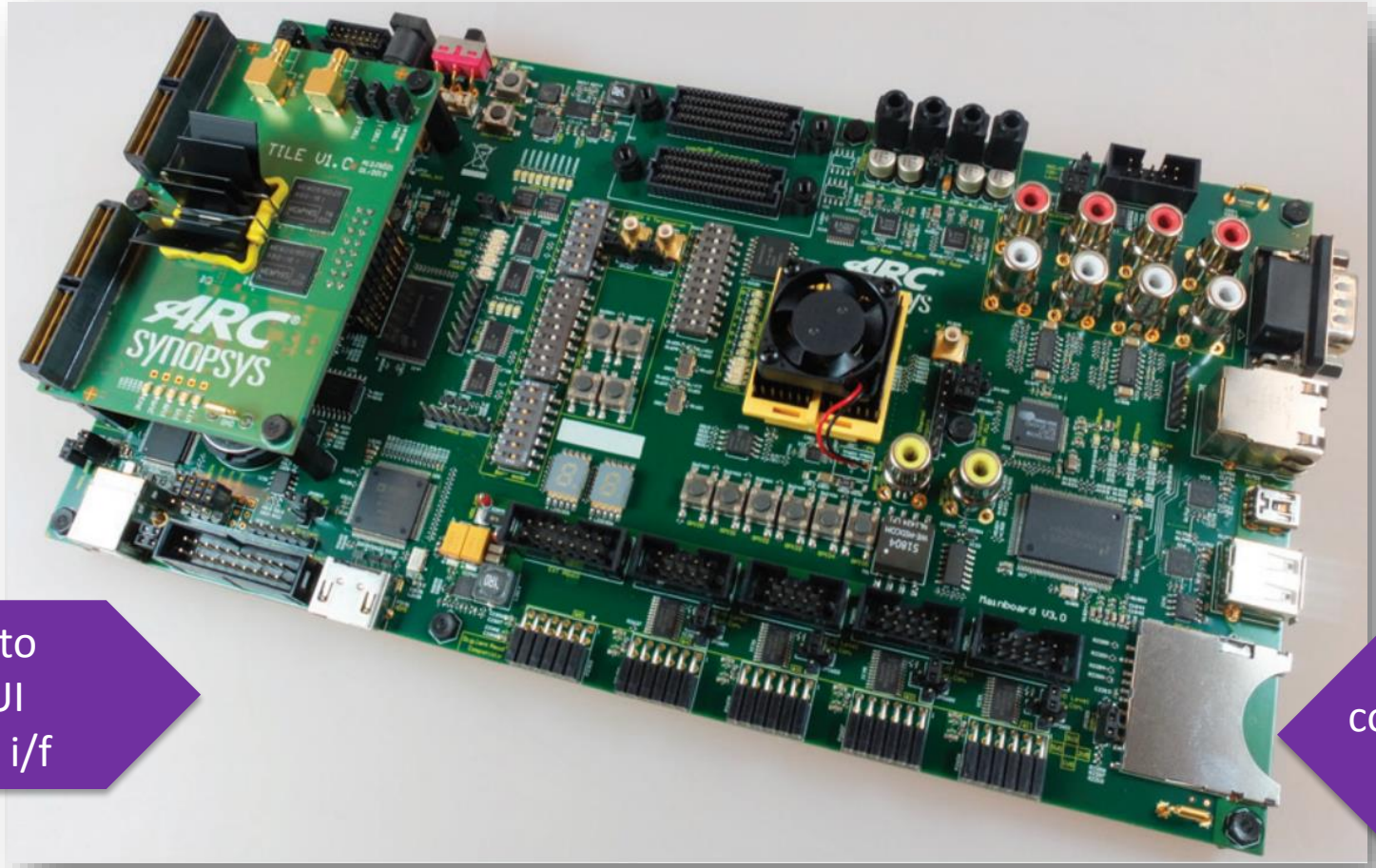


Test Setup – RC Side



Test Setup – CPU Details

ARC Software Development Platform

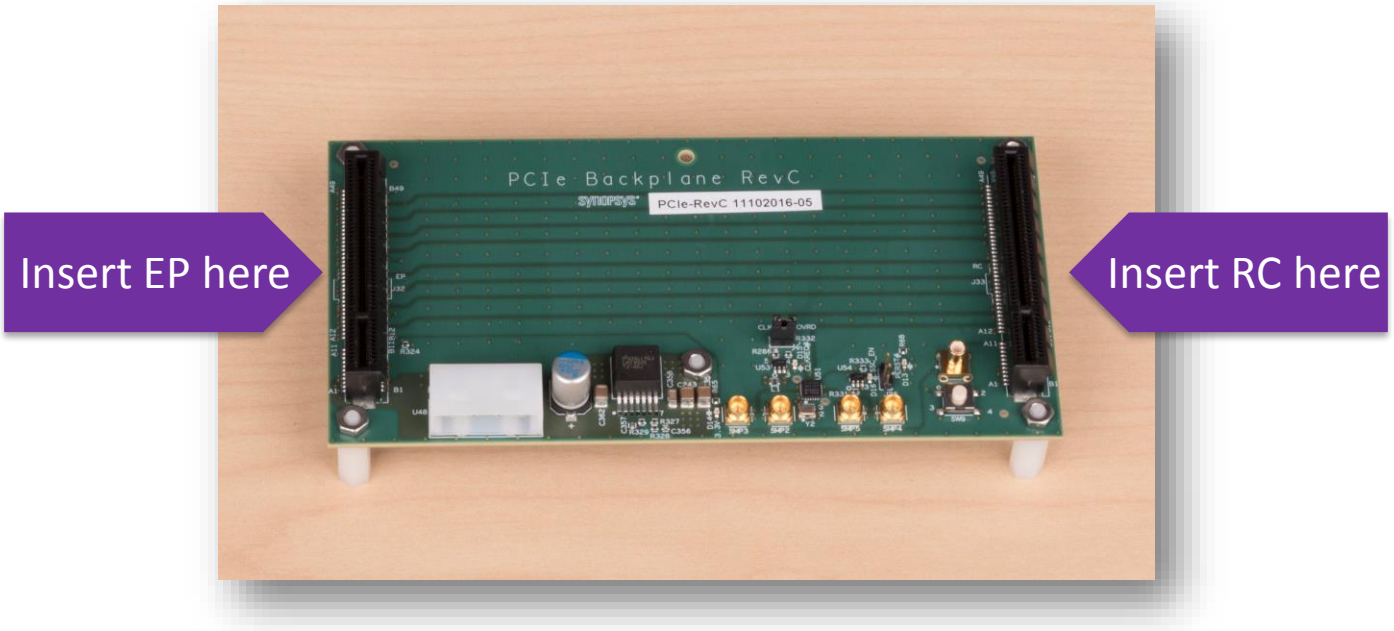


Connect to PC for GUI and shell i/f

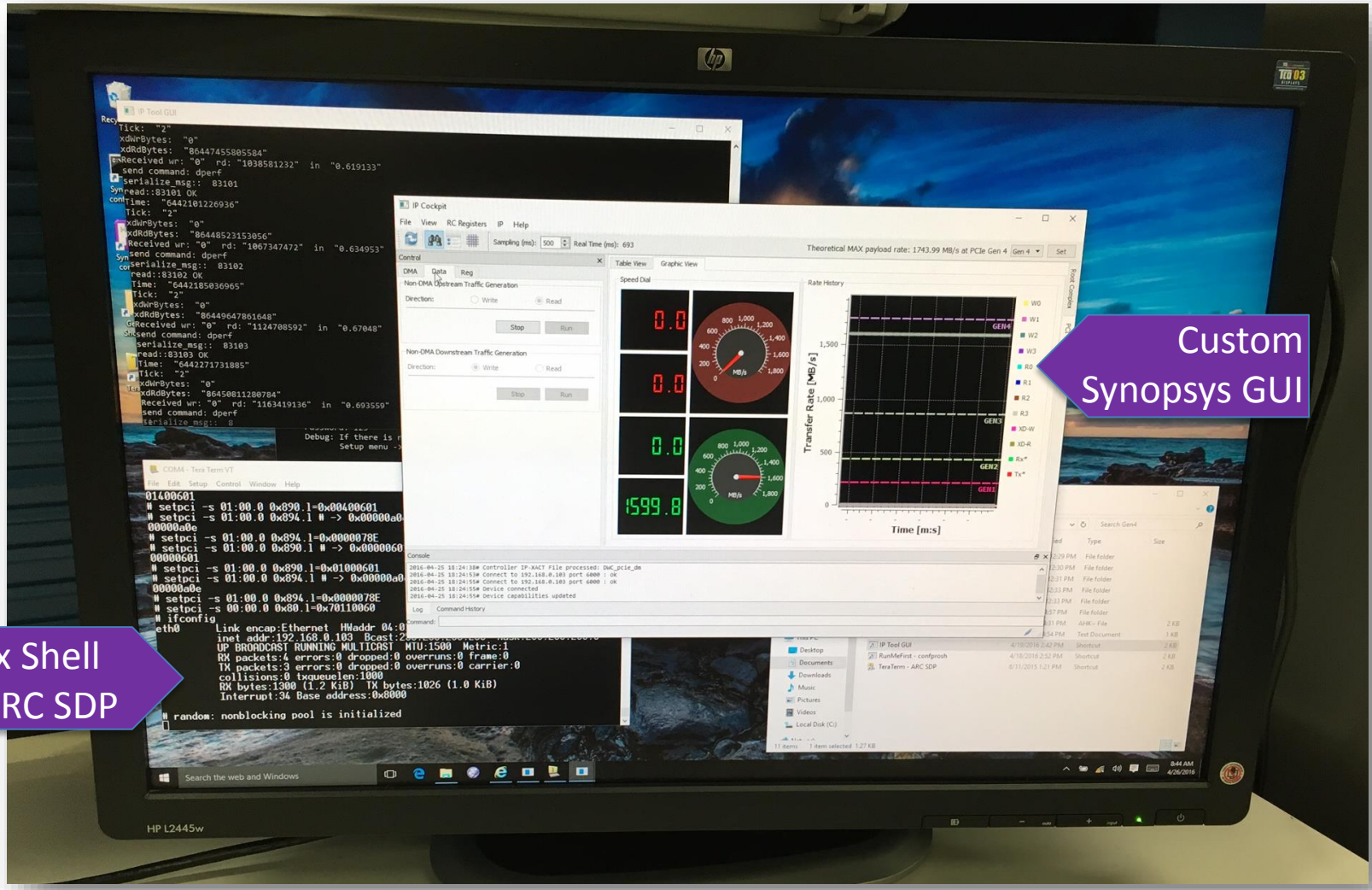
SD card containing Linux uImage

Test Setup – Backplane

Backplane connects RC IP HW to EP IP HW



Setup for Compliance Testing
Software



Testing Setup (Hardware/Software Tools)

- **PCIe 4.0 Lane Margining**
  - Support still very early even in early adopter devices
  - Device-independent eye tool
- **External protocol analyzer**
  - Interposer card reduces signal integrity
- **On Chip logic analyzer**
  - In prototyping hardware
- **Design for Reliability (RAS-DES)**
  - Full details in Richard's 2015 PCI-SIG Presentation

  https://members.pcisig.com/wg/PCI-SIG/document/8673?downloadRevision=active

Lessons From the Past…

Tracking Link Reliability: Suggested Error Counters

- PHY Layer
  - ✓ Elasticity buffer overflow/underrun
  - ✓ Symbol Errors: disparity, decoding, loss of symbol lock
  - ✓ Receiver Error
  - ✓ N_FTS Timeout
  - ✓ Framing Error
  - ✓ Deskew Error
- Data Link Layer
  - ✓ Bad packet / LCRC
  - ✓ Replays: rollovers, timeouts
  - ✓ Rx/Tx Nak DLLP
- Transaction layer
  - ✓ FC Timeout
  - ✓ Poisoned TLP
  - ✓ ECRC Error
  - ✓ UR / CA
  - ✓ Completion Timeout

PCI-SIG Developers Conference 20

Lessons From the Past…

Tracking Link Reliability: Suggested Event Counters

- PHY Layer
  - ✓ Elasticity Buffer SKP Add/Delete
  - ✓ Speed change/Link width change
  - ✓ Entry to RECOVERY state
  - ✓ L*n* state changes
- Data Link Layer
  - ✓ Tx/Rx Ack DLLP
  - ✓ Tx/Rx Update FC DLLP
  - ✓ Tx/Rx Nullified TLP
  - ✓ Rx Duplicate TLP
- Transaction layer
  - ✓ TX/RX packet types
  - ✓ FC credit exhaustion

Lessons From the Past…

Tracking Link Reliability: Suggested Status Information

- PHY Layer
  - ✓ SERDES interface signals
  - ✓ Lane detection
  - ✓ Polarity
  - ✓ SERDES power state
  - ✓ Deskew pointers
  - ✓ LTSSM states & registers
- Data Link Layer
  - ✓ Data Link Layer state
  - ✓ FC state
  - ✓ Sequence #s
  - ✓ Replay buffer depth
- Transaction layer
  - ✓ FC credit limits/usage
  - ✓ Packet queue depth

“Interesting” PCIe 4.0 Changes from 0.5 to 0.7

EIEOS Redefined for 16GT/s

- **Electrical Idle Exit Ordered Set**
- **0.5 was same as 8GT/s (eight 0s, eight 1s)**
- **0.7 changed to same frequency (sixteen each)**

Table 4-10: Electrical Idle Exit Ordered Set (EIEOS) for 8.0 GT/s and Above Data Rates

Symbol Numbers	Value	Description
0, 2, 4, 6, 8, 10, 12, 14	00h	Symbol 0: EIEOS Identifier A low frequency pattern that alternates between eight 0s and eight 1s.
1, 3, 5, 7, 9, 11, 13, 15	FFh	A low frequency pattern that alternates between eight 0s and eight 1s.

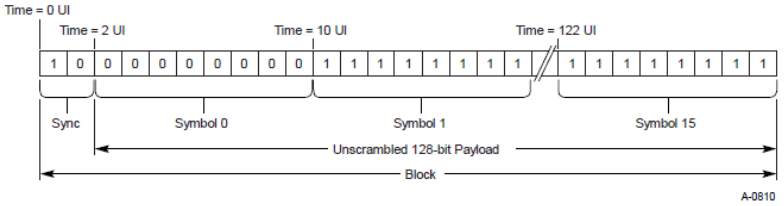
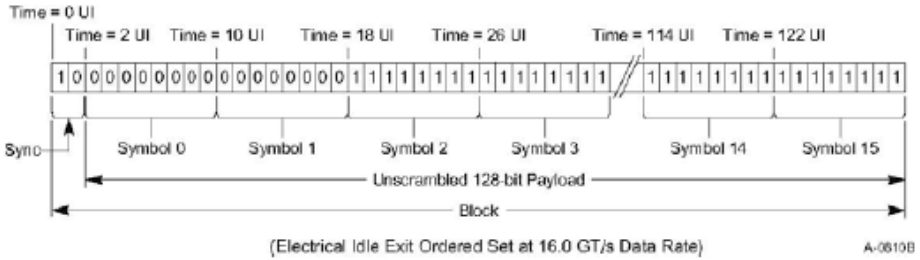


Figure 4-22: Electrical Idle Exit Ordered Set for 8.0 GT/s and Above Data Rates

Table 4-11: Electrical Idle Exit Ordered Set (EIEOS) for 16.0 GT/s Data Rate

Symbol Numbers	Value	Description
0, 1, 4, 5, 8, 9, 12, 13	00h	Symbol 0: EIEOS Identifier A low frequency pattern that alternates between sixteen 0s and sixteen 1s.
2, 3, 6, 7, 10, 11, 14, 15	FFh	A low frequency pattern that alternates between sixteen 0s and sixteen 1s.



(Electrical Idle Exit Ordered Set at 16.0 GT/s Data Rate)

- **Result: Mixed draft devices time out at 16GT/s due to never getting bit lock**

SKP Extended for 16GT/s

- **8GT/s SKP renamed to "Standard SKP" in 0.7**
- **0.7 adds "Control SKP" for Lane Margining**

Table 4-15: SKP Ordered Set with 128b/130b Encoding

Symbol Number	Value	Description
0 through (4*N - 1) [N can be 1 through 5]	AAh	SKP Symbol. Symbol 0 is the SKP Ordered Set identifier.
4*N	E1h	SKP_END Symbol. Signifies the end of the SKP Ordered Set after three more Symbols.
4*N + 1	00-FFh	(i) If LTSSM state is Polling.Compliance: AAh (ii) Else if prior block was a Data Block: Bit[7] = Data Parity Bit[6:0] = LFSR[22:16] (iii) Else: Bit[7] = ~LFSR[22] Bit[6:0] = LFSR[22:16]
4*N + 2	00-FFh	(i) If the LTSSM state is Polling.Compliance: Error_Status[7:0] (ii) Else LFSR[15:8]
4*N + 3	00-FFh	(i) If the LTSSM state is Polling.Compliance: ~Error_Status[7:0] (ii) Else: LFSR[7:0]

Table 4-17: Control SKP Ordered Set with 128b/130b Encoding

Symbol Number	Value	Description
0 through (4*N - 1) [N can be 1 through 5]	AAh	SKP Symbol. Symbol 0 is the SKP Ordered Set identifier.
4*N	78h	SKP_END_CTL Symbol. Signifies the end of the Control SKP Ordered Set after three more Symbols.
4*N + 1	00-FFh	Bit 7: Data Parity Bit 6: First Retimer Data Parity Bit 5: Second Retimer Parity Bits [4:0]: Margin CRC [4:0]
4*N + 2	00-FFh	Bit 7: Margin Parity Bits [6:0]: Refer to Section 4.2.13.1
4*N + 3	00-FFh	Bits [7:0]-: Refer to Section 4.2.13.1

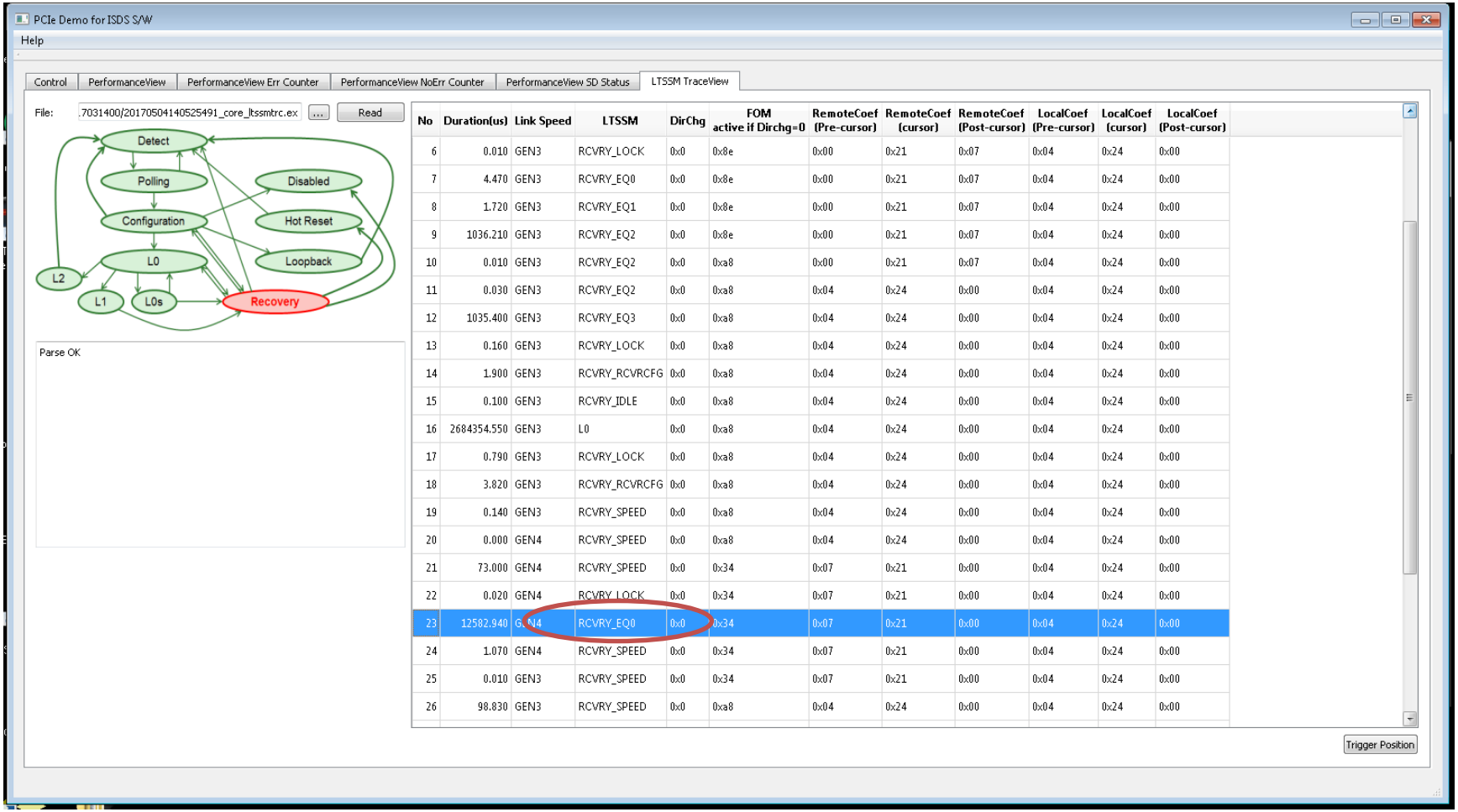
- **Result: Mixed draft devices slow due to unnecessarily looping through Recovery**

Lane Margining at the Receiver

- **Allows the system to determine how close to "the edge" each lane is operating under real conditions**
- **Relies on information from receiving PHYs – conceptual "stepping" of sampling point in eye**
- **Will be controlled by software running on the host CPU on the motherboard**
  - NOT hardware autonomous, rudimentary "protocol" spec. for s/w
  - New capability structure in Root Complex/End Point/Switch
  - Retimer control via SKP ordered sets from downstream ports
- **Result is an approximation of the eye width and optionally the eye height for each lane**

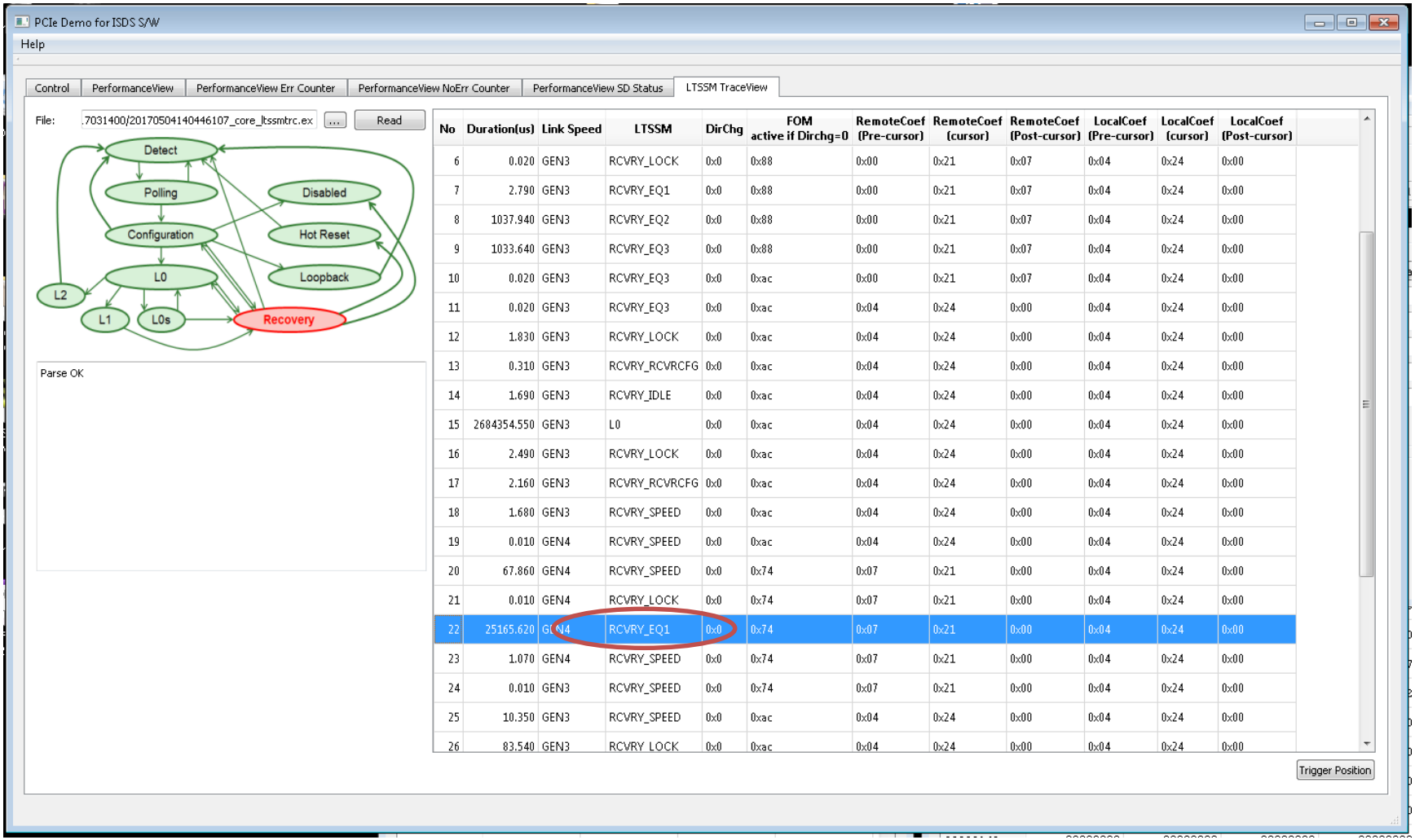
Mismatch on Implemented Specification Revisions (0.5 vs 0.7)

0.5 EP / 0.7 RC (EP Side)



No	Duration(us)	Link Speed	LTSSM	DirChg	FOM active if Dirchg=0	RemoteCoef (Pre-cursor)	RemoteCoef (cursor)	RemoteCoef (Post-cursor)	LocalCoef (Pre-cursor)	LocalCoef (cursor)	LocalCoef (Post-cursor)
6	0.010	GEN3	RCVRY_LOCK	0x0	0x8e	0x00	0x21	0x07	0x04	0x24	0x00
7	4.470	GEN3	RCVRY_EQ0	0x0	0x8e	0x00	0x21	0x07	0x04	0x24	0x00
8	1.720	GEN3	RCVRY_EQ1	0x0	0x8e	0x00	0x21	0x07	0x04	0x24	0x00
9	1036.210	GEN3	RCVRY_EQ2	0x0	0x8e	0x00	0x21	0x07	0x04	0x24	0x00
10	0.010	GEN3	RCVRY_EQ2	0x0	0xa8	0x00	0x21	0x07	0x04	0x24	0x00
11	0.030	GEN3	RCVRY_EQ2	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
12	1035.400	GEN3	RCVRY_EQ3	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
13	0.160	GEN3	RCVRY_LOCK	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
14	1.900	GEN3	RCVRY_RCVRCFG	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
15	0.100	GEN3	RCVRY_IDLE	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
16	2684354.550	GEN3	L0	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
17	0.790	GEN3	RCVRY_LOCK	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
18	3.820	GEN3	RCVRY_RCVRCFG	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
19	0.140	GEN3	RCVRY_SPEED	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
20	0.000	GEN4	RCVRY_SPEED	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00
21	73.000	GEN4	RCVRY_SPEED	0x0	0x34	0x07	0x21	0x00	0x04	0x24	0x00
22	0.020	GEN4	RCVRY_LOCK	0x0	0x34	0x07	0x21	0x00	0x04	0x24	0x00
23	12582.940	GEN4	RCVRY_EQ0	0x0	0x34	0x07	0x21	0x00	0x04	0x24	0x00
24	1.070	GEN4	RCVRY_SPEED	0x0	0x34	0x07	0x21	0x00	0x04	0x24	0x00
25	0.010	GEN3	RCVRY_SPEED	0x0	0x34	0x07	0x21	0x00	0x04	0x24	0x00
26	98.830	GEN3	RCVRY_SPEED	0x0	0xa8	0x04	0x24	0x00	0x04	0x24	0x00

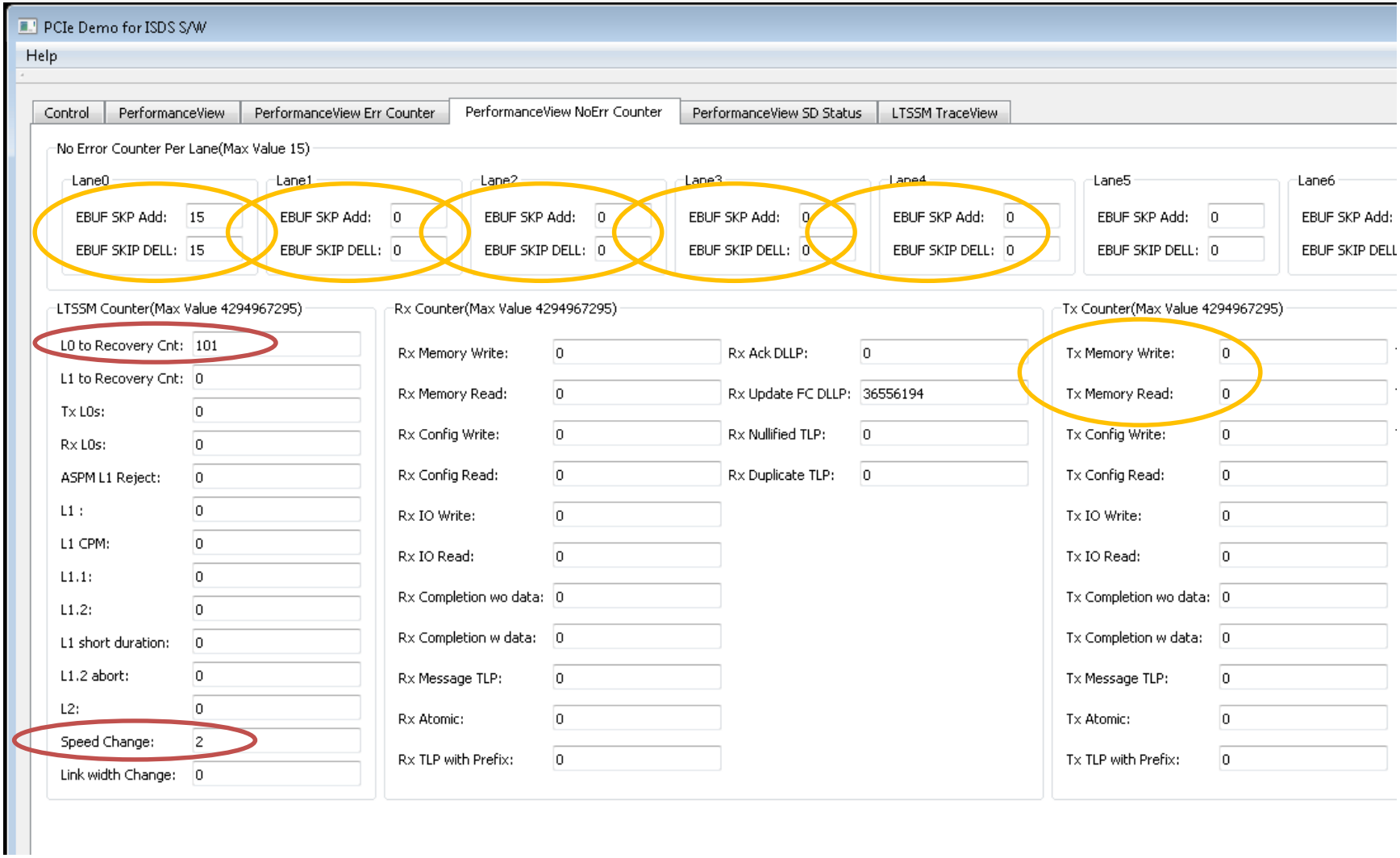
0.5 EP / 0.7 RC (RC Side)



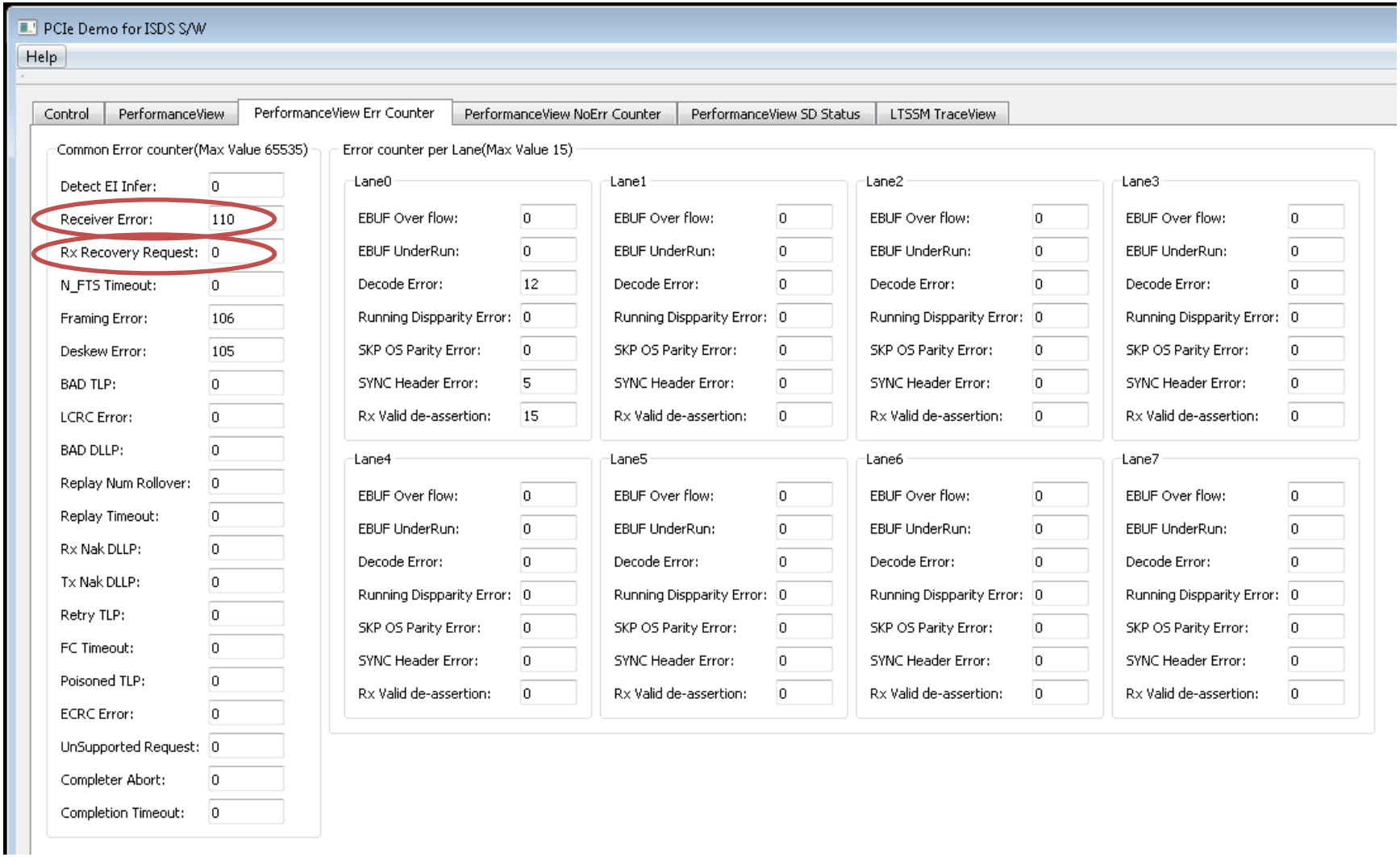
No	Duration(us)	Link Speed	LTSSM	DirChg	FOM active if Dirchg=0	RemoteCoef (Pre-cursor)	RemoteCoef (cursor)	RemoteCoef (Post-cursor)	LocalCoef (Pre-cursor)	LocalCoef (cursor)	LocalCoef (Post-cursor)
6	0.020	GEN3	RCVRY_LOCK	0x0	0x88	0x00	0x21	0x07	0x04	0x24	0x00
7	2.790	GEN3	RCVRY_EQ1	0x0	0x88	0x00	0x21	0x07	0x04	0x24	0x00
8	1037.940	GEN3	RCVRY_EQ2	0x0	0x88	0x00	0x21	0x07	0x04	0x24	0x00
9	1033.640	GEN3	RCVRY_EQ3	0x0	0x88	0x00	0x21	0x07	0x04	0x24	0x00
10	0.020	GEN3	RCVRY_EQ3	0x0	0xac	0x00	0x21	0x07	0x04	0x24	0x00
11	0.020	GEN3	RCVRY_EQ3	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
12	1.830	GEN3	RCVRY_LOCK	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
13	0.310	GEN3	RCVRY_RCVRCFG	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
14	1.690	GEN3	RCVRY_IDLE	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
15	2684354.550	GEN3	L0	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
16	2.490	GEN3	RCVRY_LOCK	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
17	2.160	GEN3	RCVRY_RCVRCFG	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
18	1.680	GEN3	RCVRY_SPEED	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
19	0.010	GEN4	RCVRY_SPEED	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
20	67.860	GEN4	RCVRY_SPEED	0x0	0x74	0x07	0x21	0x00	0x04	0x24	0x00
21	0.010	GEN4	RCVRY_LOCK	0x0	0x74	0x07	0x21	0x00	0x04	0x24	0x00
22	25165.620	GEN4	RCVRY_EQ1	0x0	0x74	0x07	0x21	0x00	0x04	0x24	0x00
23	1.070	GEN4	RCVRY_SPEED	0x0	0x74	0x07	0x21	0x00	0x04	0x24	0x00
24	0.010	GEN3	RCVRY_SPEED	0x0	0x74	0x07	0x21	0x00	0x04	0x24	0x00
25	10.350	GEN3	RCVRY_SPEED	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00
26	83.540	GEN3	RCVRY_LOCK	0x0	0xac	0x04	0x24	0x00	0x04	0x24	0x00

Signal Integrity Issues

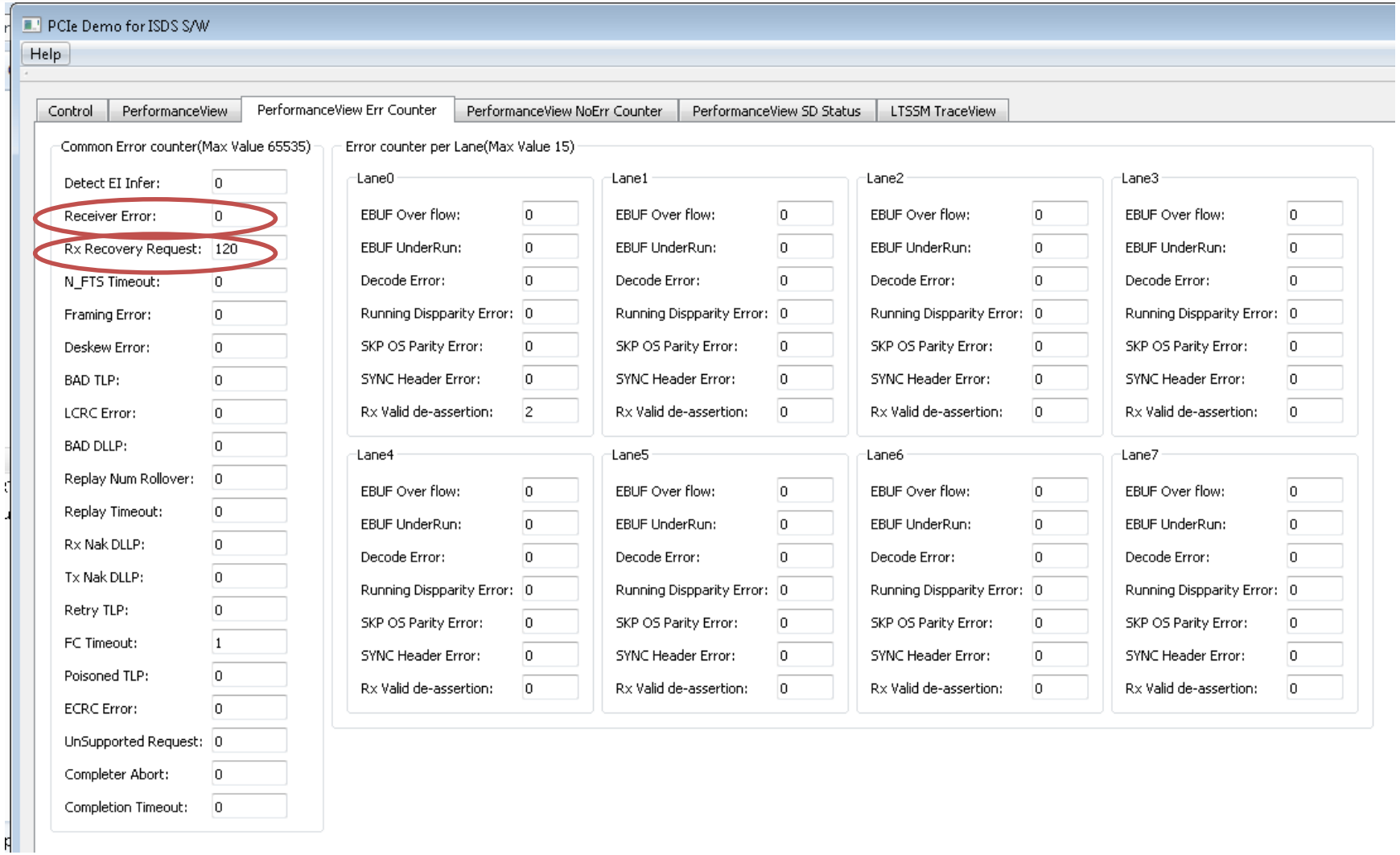
Unclean Link (RC Side)



Unclean Link (RC Side)

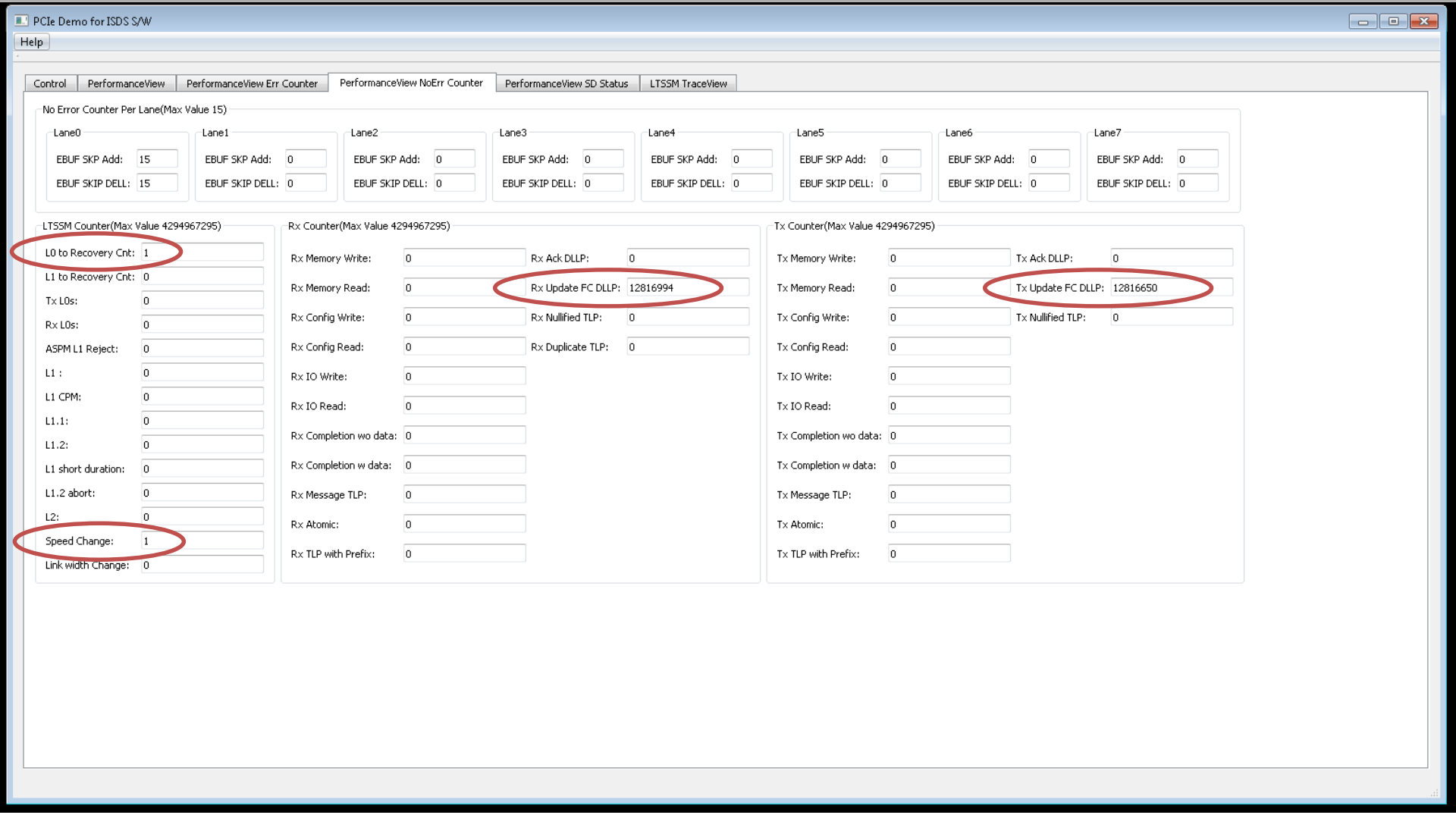


Unclean Link (EP Side)

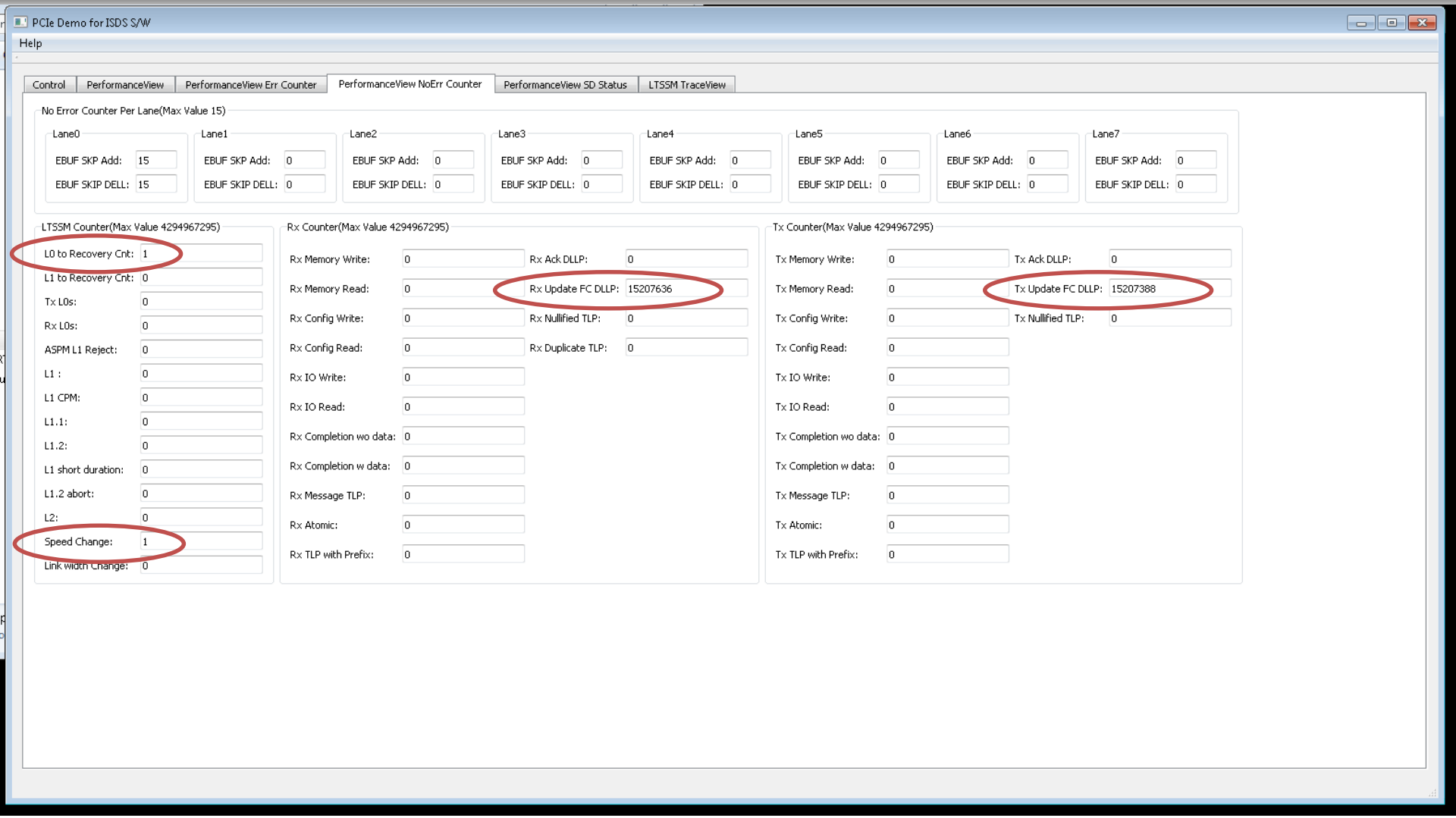


Clean 16GT/s Link Up

Clean Link-up



Clean Link-up (Time Elapsed)



Wrap-up

Failure Manifestations

- **Link retrains on random interval**
  - Signal integrity
- **Link retrains on a regular interval**
  - Points to protocol problem – like missing SKPs, etc.
- **Other error indications**
  - Widely differing SKP add and delete across lanes
  - Multiple speed changes
  - Packet transmission counts not increasing
- **Frequency of errors**
  - Low frequency less concerning than high

Resolutions

- **Try forcing presets to eliminate link negotiation problems**
- **Force maximum link speed lower to see if problems remain at lower speeds**
- **Limit number of lanes (tape, interposer, etc.)**
- **Get a hardware analyzer in-system**

Conclusions

- **Look at details behind link retraining for clues to “mixed draft” links**
  - Complete failure to link at 16GT/s
  - Periodic retraining while at 16GT/s
- **Design to the latest draft**
  - 0.7 is solid
  - 0.9 as soon as it’s published
- **Build in design-for-reliability / design-for-debug features to assist in link bring-up**

Thank you for attending the PCI-SIG Developers Conference 2017.

For more information please go to www.pcisig.com